View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DigitalCommons@USU

Analyzing sea-level change on the east coast with spatiotemporally correlated data

David W. Coats, Candace Berrett, William F. Christensen, Nathan Sandholtz

Abstract

Increasing rates in sea-level rise imply drastic consequences for U.S. coastal populations, infrastructure, ecological systems, and natural resources in the coming decades. These direct impacts will lead to negative repercussions in public health, biodiversity, tourism, and other aspects of the global economy. Using hourly tide readings from the past 30 years at 38 gauges along the east coast, we wish to develop a model that will allow us to analyze the trends in this type of data and to accurately and precisely predict sea-level change along the east coast. The model developed is an iterative generalized additive model that will use spatial and temporal dependence between gauges and across time, allowing us to predict sea-level change all along the east coast, not only at the stations for which we have data. Here, the methodology and components of our current model will be discussed as well as an overview of results. We will also address the model's shortcomings and the work that is currently being done to improve the accuracy and efficiency of its predictions.

Introduction

The Intergovernmental Panel on Climate Change (IPCC) estimates that the global sea level is currently rising at a rate of 3 millimeters per year and this rate is expected to increase over the coming century [1]. This increase in the rate of sea-level rise could lead to changes that will affect many aspects of daily life and the global economy, thus accurate predictions and thorough understanding of the trends in this process are vital for preparation for these changes. The data used in this study comes from NCAR (National Center of Atmospheric Research) and consists of tide gauge readings taken hourly from 1979 to 2009 from 38 stations along the east coast of the United States. Tide gauges are instruments deployed at coastal sites around the world that directly measure sea-level change as compared to a determined base level. For the 38 tide gauges in this study, sea-level change is measured as deviation from the mean high water level for that station over a 19 year epoch.

http://www.oco.noaa.gov/tideGauges.html

Figure 1: Example of a tide gauge from NOAA (National Oceanic and Atmospheric Administration)

The 38 stations where these tide gauges are located range in location from Bar Pilots Dock–St. Johns River in Florida to East Port–Passamaquoddy Bay in Maine. At some of these stations there are many missing observations due either to malfunctions in the tide gauge or because at that time there was no gauge in that location. These missing values lead to complications in modeling and predicting, but these issues will be addressed later on.

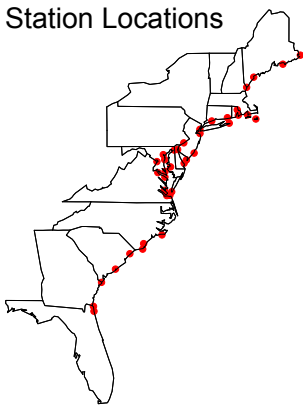

Figure 2: Locations of the 38 tide gauges along the east coast

Due to factors dealing with global location, sea-level change is not constant across space in that it is different depending on location. An important observation when considering spatial data such as this is that sites that are closer together are more likely to be closely related than sites that are further part. Our model will take into account these spatial correlations, and this theory will eventually allow us to predict up and down the coast because of the spatial relationship between these sites and sea-level change.

Exploratory Analysis: Exposing Trends

In order to understand the temporal trends in our data, we identify patterns in sea-level change at each station over time, and model these trends for individual sites. After identifying and exploring these trends in individual sites we will seek to combine these trends in an iterative spatial generalized additive model that will use correlation between sites in order to predict at any location along the east coast. Because ocean tides are greatly influenced by the moon and its cycles, we averaged the 30 years of data by lunar months which are approximately 28 days long, resulting in 371 lunar month averages for each station. A general linear trend can be seen in these lunar month averages over time in all of the stations, but there are other trends going on in addition to a simple linear relationship. The graphic below shows the lunar month averages across time with a simple linear fit plotted on top (the blue line). Note the missing data between 1996 and 1998.

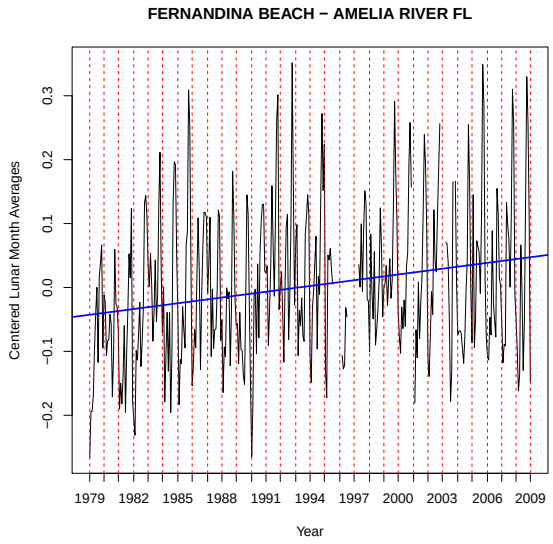

Figure 3: The linear trend in lunar month average across time seen at a single station

One can see the linear trend in this plot, but there appears to be a cyclical effect within years that could be some type of seasonal trend. After accounting for the simple linear trend at each station, we compute the residuals and see a definite pattern. The residuals for the southern most stations have what we will call an M-curve, but as we move north, station by station this M-shape appears to flatten out into flatter, more unimodal curves suggesting that this M-curve effect changes across space. We fit these curves individually for each station with B-splines. A B-spline with k knots splits the covariate space into $k+1$ regions fitting a cubic polynomial to the data in each region. The spline is constrained so that the polynomials are differentiable at the knot points resulting in a smooth non-linear fit to the data. The figure below shows the residuals for a north station and a more southern station fit with B-splines with 6 knots. Red dots are observations from later years, and the yellow are from earlier years.

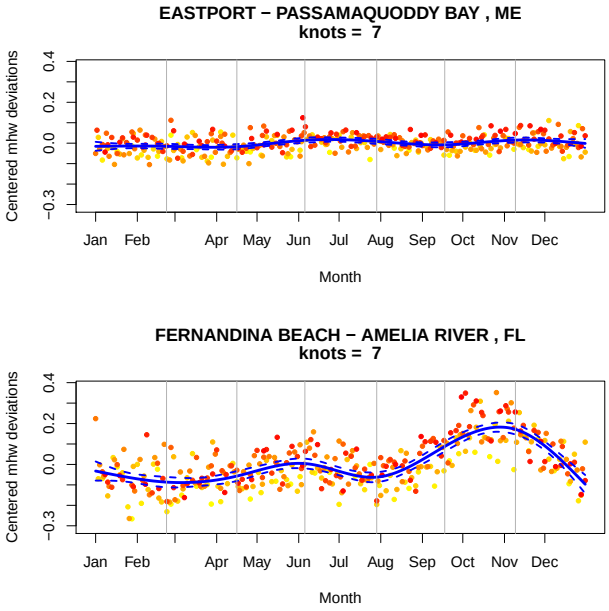

Figure 4: The seasonal trends for a northern and southern station fit with B-splines with 6 knots

These splines appear to be good fits to the residuals and because the seasonal trend seems to have spatial correlation, we will attempt to fit spatial terms to these splines so that we can model the relationship of this trend with spatial location. Now that we have uncovered some important trends in the data, we will develop a model that iteritively fits these trends and uses spatial correlation to explain the relationships between sites allowing for more extensive predictive power and better understanding of sea-level changes across space and time.

Iteritive Generalized Additive Model

Generalized additive models are a generalization of linear models in which the predictions depend on smooth functions of the covariates [2]. The response variable Y follows an exponential family distribution, which in our case will be $\mathcal{N}(0, \sigma^2)$. We will model Y_{it}, sea-level change at station i and time t, with an intercept μ_i that will be the overall mean of each station, a linear term, and a spline term that accounts for the seasonal trend. This model will be expressed in the following manner:

$$Y_{it} = \mu_i + t\beta_i + \sum_{j=1}^{k+1} f_j(t^*)\xi_{ij} + \epsilon_{it} \qquad \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

where $t\beta_i$ is time centered and scaled times the coefficient vector β for station i, $f_j(t^*)$ is the j^{th} polynomial fit to covariate region $(k+1)$ where $k = 6$ is the number of knots in our B-spline, t^* is the day of the year, ξ_{ij} is the coefficient fit to the j^{th} covariate region for the i^{th} station, and ϵ_{it} is the unexplained variance in sea-level change at station i and time t.

We refer to our model as an iteritive generalized additive model because we will fit the terms iteritively. We will fit the linear term to the residuals of the model containing only μ_i, then we will fit the spline term to the residuals of the model with the intercept plus the linear term. We fit the model in this manner in order to address the problem we have with missing observations. The missing observations were all replaced with the overall mean of the station to which they belong as a beginning value. For each step in fitting the model we predict Y_{it} and then replace the formerly missing values with $\hat{Y}_{it}$ iteritively until the predictions of these observations and the coefficients of the model converge to specific values. Fitting the model this way allows us to update the values of the missing observations based on the trends in the data step by step resulting in better estimates for these values at every step and a better final estimate.

Inclusion of Spatial Correlation in Model

Previously we fit a linear trend and a spline term individually to each station; now we will model the relationship or correlation of these trends from station to station. Understanding how spatial distance affects correlation between stations will allow us to be able to make inference along the coast between our stations. Given the data at our stations and the distances from new locations to our stations, we will be able to predict sea-level changes at these new locations.

A variogram is a function describing the correlation between points that are different distances apart. Variograms can be modeled with different spatial correlation structures that behave differently depending on how your data is spatially correlated. The Matern, exponential, Gaussian, and spherical functions are examples of spatial correlation structures; by exploring the fits of these different functions to the residuals left over after taking out the linear trend, we decide that the spherical function is the best fit. Assuming the spherical spatial correlation structure is a good fit to the variogram of our data, the correlation between two observations with distance $r < \phi$ between them is

$$c(r) = (1-n)\frac{3r}{2\phi} + \frac{1}{2}\left(\frac{r}{\phi}\right)^3$$

for all observations for which $r > 0$ where ϕ is the range over which the correlations will be nonzero, and n is the nugget. The range ϕ refers to the distance at which the variogram appears to level out because points with distances greater than ϕ are not correlated. For reference to the terminology used in modeling variograms, the semivariance at ϕ is σ^2, referred to as the sill, and the partial sill is the (sill - nugget); nugget is the semivariance at distance = 0, meaning that if the nugget is non-zero, that there is variance among points that are very close together, indicating underlying trends in the residuals.

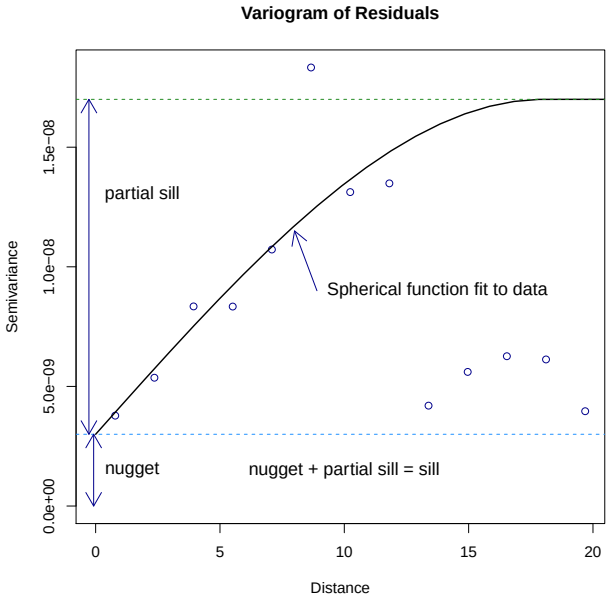

Figure 5: Variogram of the residuals from the linear model fit with spherical correlation function

In order to estimate the coefficient β_i for the linear term, we first fit $Y_{it} - \mu_i = t\beta_i^*$,where β_i^* is the coefficient fit to centered time t for station i. We then estimate the spatial covariance matrix $\Sigma_\beta(\theta)$ of β_i^* with the spherical corellation function where $\theta = (\sigma^2, \phi, \tau^2)$ with σ^2 = the sill, ϕ = the range parameter, and τ^2 = the nugget effect. We then fit $\beta_i = \alpha_0 + x_i\alpha_1 + x_i^2\alpha_2$ using generalized least squares with $W{=}\Sigma_\beta(\theta)$ being the variance matrix such that the coefficient vector $\alpha = (X'W^{-1}X)^{-1}X'W^{-1}\beta^*$, and x_i being the spatial location of site i, and $\alpha = (\alpha_0, \alpha_1, \alpha_2)$. Using these methods, we have modeled the trend of how the linear effect changes across space while maintaining accurate modeling at individual stations.

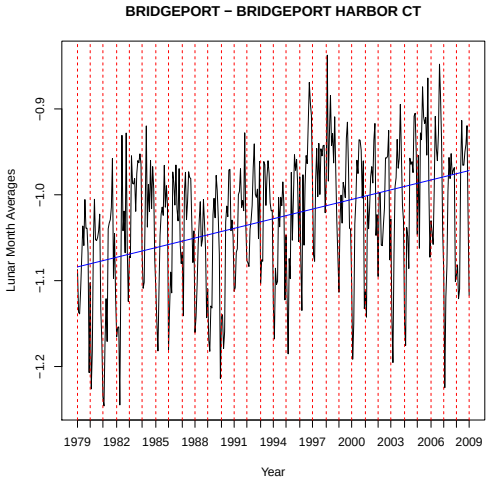

Figure 6: The linear trend in lunar month average across time seen at a single station

Now for the spline term $\sum_{j=1}^{k+1} f_j(t^*)\xi_{ij}$, we model $\xi_i j$ in the same manner as we modeled β_i but fit to the residuals of the model including $t\beta_i$. After estimating $V_j = \Sigma_\xi(\theta)$, ξ_{ij} for the i^{th} station and j^{th} of $k+1$ regions across covariate space, and solving for $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \gamma_{j2}) = (X'V_j^{-1}X)^{-1}X'V_j^{-1}\xi^*$, we have $\xi_{ij} = \gamma_{j0} + x_i\gamma_{j1} + x_i^2\gamma_{j2}$ by generalized least squares with variance matrix V_j.

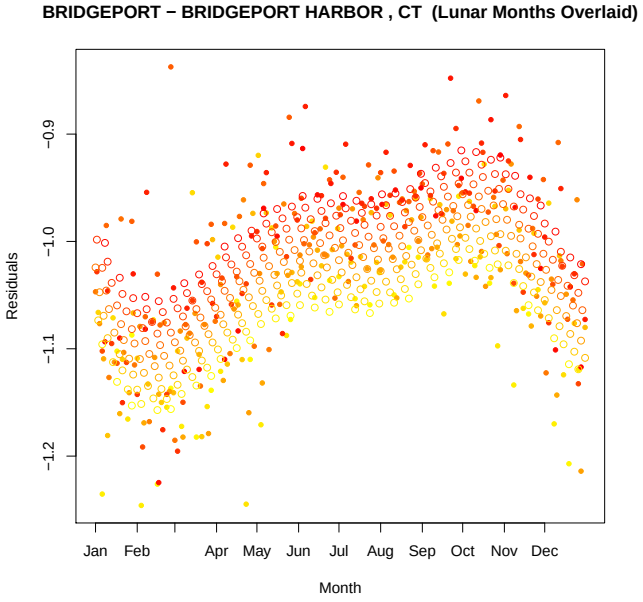

Figure 7: The linear trend in lunar month average across time seen at a single station

In the figure above we see the predicted values from our model as hollow points and residuals from the previous state of our model as solid points with yellow indicating earlier years, and red indicating more recent years. Because of the positive linear trend modeled by $t\beta_i$, the model predicts the m-curves for more recent years to be up higher than the m-curves for earlier years. Because the splines are constrained to be the same shape across years, the shape is flattend out more than it should be, making prediction of the more extreme points very difficult. In future work we will explore fitting yearly means to our model in an attempt to increase the accuracy of our model. In the following plot we can see this trend in the lunar month averages compared to our predicted lunar month averages (blue); our predictions look good, but it appears that if our curve was magnified, it would predict much better. We believe that accounting for the yearly fluctuation in the m-curves will allow the model to be more flexible in reaching the data from more extreme years.

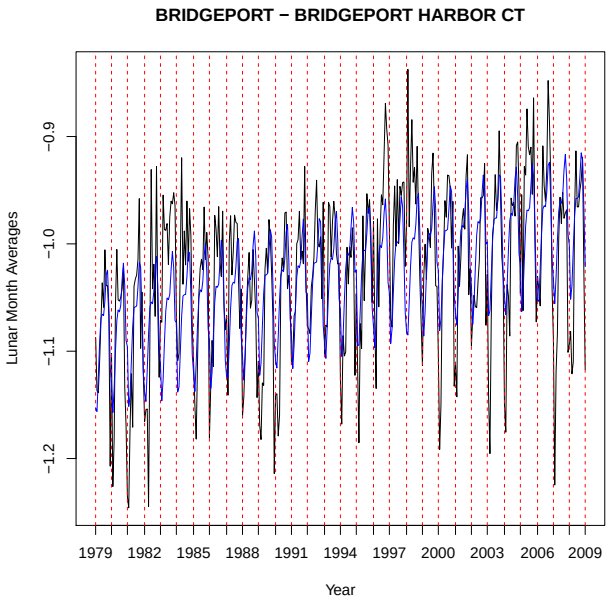

Figure 8: The linear trend in lunar month average across time seen at a single station

Once we fit the model to take out the yearly trend mentioned previously, we aim to model the remaining residuals by extracting eigen vectors in an effort to account for latent factors in the data. Not only will applying this type of latent variable model allow our predictions to be more accurate, it will shed light on possible variables that affect sea-level change that we did not discover in our analysis.

Conclusion

Modeling sea-level change with an iterative generalized additive model that accounts for spatial and temporal correlation between observations results in a versatile model that can be used as a strong predictive tool across a wide variety of locations. The change in rate of the rise in sea-level cannot by modeled as a positive simple linear model because of its complex relationships with time and space. By applying spatial correlation to the model accounting for the overall linear trends and seasonal trends by station, we have good predictive power, but there are trends in the data that warrant further investigation. It was mentioned earlier that the uncertainty of our predictions would be distributed $\mathcal{N} \sim (0, \sigma^2)$. After accounting for the trends mentioned in the last section, we will estimate the uncertainities of our model or $\hat{\sigma}^2$, and evaluate the model's efficiency.

References

[1] 2007 IPCC Fourth Assisment Report, Climate Change 2007: Working Group 1: The Physical Science Basis, Chapter 5.5 - Changes in Sea Level

[2] Hastie, T.J. and Tibshirani, R.J. 1990. *Generalized Additive Models.* Chapman & Hall/CRC.